**#Probabilità**

**#Binomiale**

#P ➔ somma delle probabilità di tutti i possibili esiti (0, 1, 2, n)

pbinom(n\_successi, n\_tentativi, probabilità)

#1-pbinom ➔ successi maggiori di n

pbinom(n\_successi, n\_tentativi, probabilità, lower.tail = FALSE)

#D ➔ probabilità di n\_successi esatti

dbinom(n\_successi, n\_tentativi, probabilità)

#**Geometrica** **→** numero di fallimenti PRIMA del successo   
#P ➔ somma delle probabilità di tutti i possibili insuccessi

pgeom(n\_insuccessi, probabilità)

#1-pgeom ➔ insuccessi maggiori di n

pgeom(n\_insuccessi, probabilità, lower.tail = FALSE)

#D ➔ probabilità di n\_insuccessi esatti

dgeom(n\_insuccessi, probabilità)

**#Ipergeometrica ➔** estrazione di palline senza reimbussolamento#P ➔ somma dei possibili esiti delle estrazioni

phyper(n\_successi, n\_favoverevoli, n\_sfavorevoli, n\_tentativi)

#1-phyper➔ successi superiori a n\_successi

phyper(n\_successi, n\_favoverevoli, n\_sfavorevoli, n\_tentativi, lower.tail = FALSE)

#D ➔ probabilità di n\_successi esatti

dhyper(n\_successi, n\_favoverevoli, n\_sfavorevoli, n\_tentativi)

**#Poisson ➔** prove ripetute con probabilità molto piccola (dato dal testo)

#P ➔ somma dei possibili esiti delle per i successi

ppois(n\_successi, n\_eventi\_medio)

#1-ppois ➔ successi superiori a n\_successi (>) (se >= n\_succ – 1)

ppois(n\_successi, n\_eventi\_medio, lower.tail = FALSE)

#D➔ probabilità di n\_successi esatti

dpois(n\_successi, n\_eventi\_medio)

**#Normale**

#P ➔ valore sotto la curva tra valore minimo della curva e il valore indicato

pnorm(valore, mean = media, sd = deviazione\_standard)

#1-pnorm ➔ valore sotto la curva tra il valore indicato e il valore massimo

pnorm(valore, mean = media, sd = deviazione\_standard, lower.tail = FALSE)

#dnorm ➔ valore della probabilità al punto indicato

dnorm(valore, mean = media, sd = deviazione\_standard)

#Q➔valore del quantile (valore per cui n% è minore di tale valore) (se maggiore, 1-n) (inversa della distribuzione

qnorm(valore percentuale richiesto, mean = media, sd = deviazione\_standard)

**#Uniforme**

#P ➔ valore sotto la curva tra valore minimo e il valore indicato

punif(valore, min = a, max = b)

#1-punif ➔ valore sotto la curva tra il valore indicato e il valore massimo

punif(valore, min = a, max = b, lower.tail = FALSE

#D ➔ valore della probabilità al punto indicato

dnorm(valore, min = a, max = b)

**#Esponenziale** (lamba è sull’asse y, media = 1/lamba)

#P ➔ valore sotto la curva tra valore minimo e il valore indicato

pexp(valore, rate = 1/media)

#1-pexp ➔ valore sotto la curva tra il valore indicato e il valore massimo

pexp(valore, rate = 1/media, lower.tail = FALSE)

#D ➔ valore della probabilità al punto indicato

dexp(valore, rate = 1/media)

#Q➔valore del quantile (valore per cui n% è minore di tale valore) (se maggiore, 1-n)data una probabilità restituisce il numero che rappresenta nella funzione quella probabilità, cioè l’area sottesa dalla curva.

qexp(valore percentuale richiesto, rate = 1/media)

proprietà assenza di memoria: P(X>M | X>N) = P(X>M-N)

##################################

**#Statistica**

data("InsectSprays") #selezione dataset

str(InsectSprays) #anteprima dataset

datiWithoutB = subset(dati, dati$group != "2") #rimozione dato gruppo (seleziona solo i dati con gruppo diverso da 2)

newdata <- InsectSprays$count[InsectSprays$spray != "A"] #considero solo un gruppo (da count stai selezionando solo quelli che non sono associati allo spray A.)

summary(InsectSprays$count) #sommario dataset

quantile(datiwithoutB$var, 0.6, na.rm = TRUE) #solo il 60% dei dati ha var inferiore, se 60% var sup, 0.4 al posto di 0.6

(boxplot(InsectSprays$count) #grafico

abline(h = 20, col = "Blue") #visualizzazione valori outlier

hist(InsectSprays$count) #istogramma

mean(InsectSprays$count[InsectSprays$spray == "A"]) #media

median(InsectSprays$count[InsectSprays$spray == "A"]) #mediana

t.test(A$var1, A$var2, alternative = "two.sided/greater/less", mu = media, paired = TRUE conf.level = 0.05) #test due campioni confidenza 0.05 (e due variabili non indipendenti il valore medio di una distribuzione si avvicina ad un valore di riferimento?)

t.test(A$var1, alternative = "two.sided/greater/less", mu = media, conf.level = 0.05) #test un campione confidenza 0.05

binom.test(c(N successi, N fallimenti), alternative = "two.sided/greater/less", ipotesi nulla = value)

binom.test(vettore lunghezza 2(numero successi e fallimenti), conf.level=0.95)

c(5,95) -> vettore lungo 2, 5 successi e 95 fallimenti

binom.test(c(5,95), conf.level = 0.95)

ATTENZIONE: PARTE POCO CHIARA, VAI A VEDERE GLI ESEMPI

###########

#distribuzione-> media, mediana

media>mediana -> asimmetrica con valori più piccoli

mediana>media -> asimmetrica con valori più grandi

media = mediana -> all’incirca è simmetrica

varianza->sd(dataset)^2

#PER CONFRONTARE ELEMENTI IN MANIERA QUALITATIVA

Il trattamento insetticida A sembra eliminare gli insetti in maniera simile al trattamento C

boxplot(count~spray, data = InsectSprays)

#MEDIA SOLO SU UN ELEMENTO

newdataA <- InsectSprays$count[InsectSprays$spray == "A"]

mean(newdataA)

datiWithoutB = subset(dati, dati$group != "2") #eliminare elemento

table(InsectSprays$spray, useNA="always") #per vedere tutti gli elementi con somme

prima boxplot, poi abline(h=valore rischiesto, col = “red”) #outlier

#rinomino

A<-InsectSprays$count[InsectSprays$spray=='A']

sum(table(A)) o length(A) #somma le categorie

#contare elementi nulli

na<-InsectSprays$count[InsectSprays$spray=='na'] #per elementi vuoti

length(na)

oppure

table(InsectSprays$spray.useNa=="always")

#MEDIA CAMPIONARIA, ricordarsi il na.rm=true

#ipotesi alternativa

○ pvalue > significatività  
a favore dell’abbandono ipotesi alternativa e a favore ipotesi nulla [o nessuna delle precedenti o stessa media o differenza delle medie = 0 ->risposte esercitazioni]

○ p value < significatività a favore ipotesi alternativa e all’abbandono ipotesi nulla [o gruppo 1 fa meno del gruppo 2?] [media x != media y]

se osservazioni sono <30 devi ipotizzare la normalità, se sono >30 non aggiungo altro

##################################

**#Probabilità**

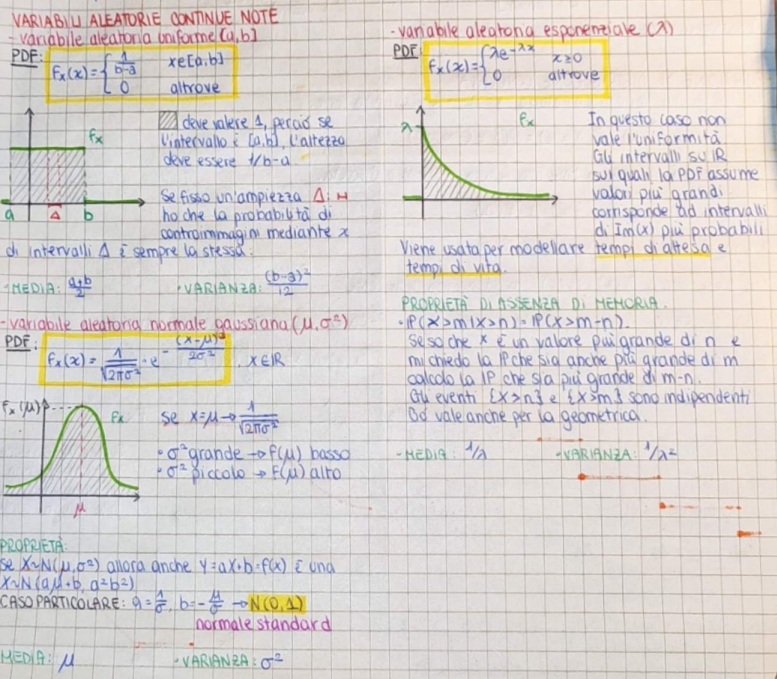
P(A ∪ B) = P(A) + P(B) − P(A ∩ B) (intersezione può essere uguale a 0 (esce 1 e esce 2 non hanno intersezioni))

P(A ∩ B) = P(A) \* P(B) #eventi **indipendenti**

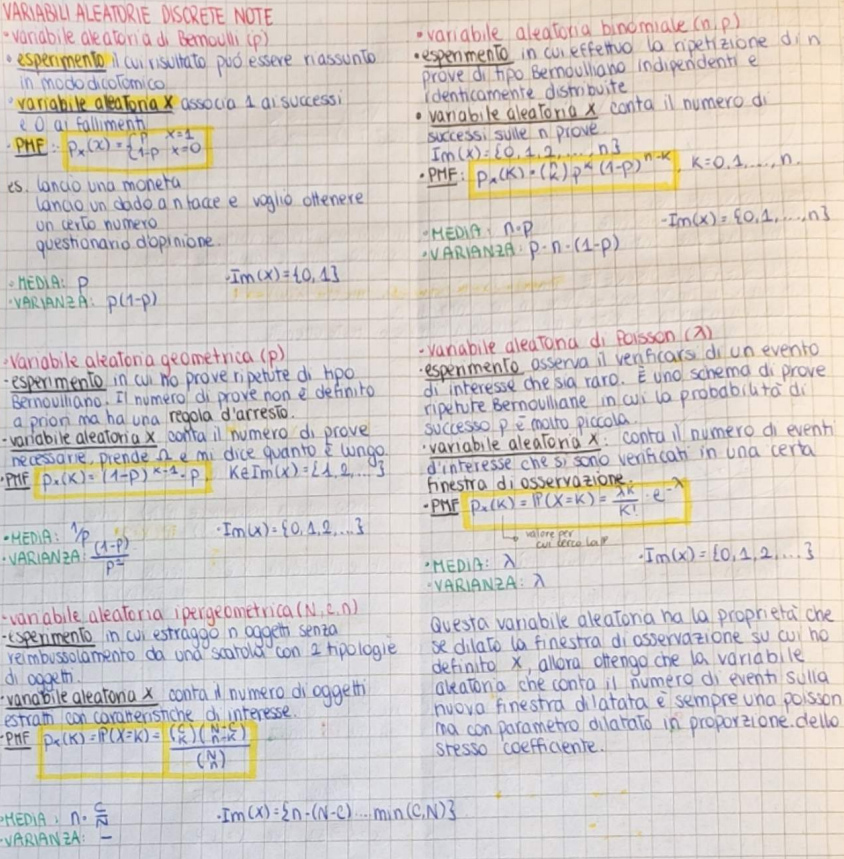
P(A ∩ B) = P(A) \* P(B|A) = P(B) \* P(A|B) #eventi **dipendenti**

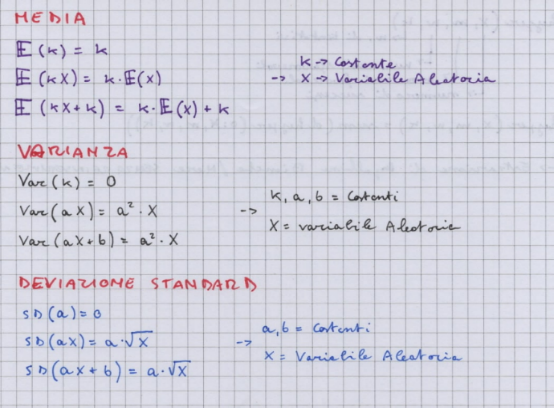
P(∩) = 1 - P(A∪B)

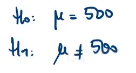
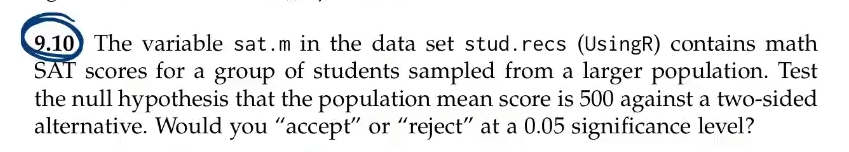
P(A) = P(A|B) \* P(B) + P(A|) \* P()

P(A|B) = P(B ∩ A)/P(A) #Bayes

= #DeMorgan  
 = #DeMorgan





#t.test e binom.test: esempi  


t.test(stud.recs$sat.m,alternative = "two.sided",mu = 500)

One Sample t-test

data: stud.recs$sat.m

t = -2.5731, df = 159, p-value = 0.01099

alternative hypothesis: true mean is not equal to 500

95 percent confidence interval:

475.1437 496.7313

sample estimates:

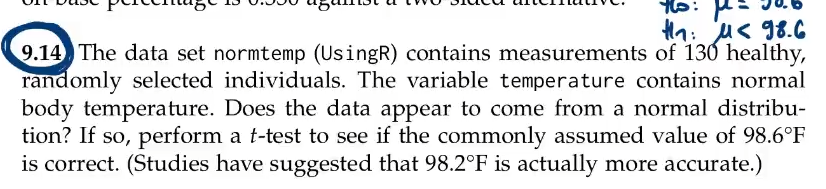
mean of x

485.937

0.05>0.01099? sì! Allora rifiuto ipotesi nulla in favore dell’alternativa

Two.sided->mi chiede se è diverso, non < o >

475.1437 496.7313 sono tutti valori plausibili, infatti 500 non c’è->alternativa



t.test(normtemp$temperature, alternative = "less", mu = 98.6)

One Sample t-test

data: normtemp$temperature

t = -5.4548, df = 129, p-value = 1.205e-07

alternative hypothesis: true mean is less than 98.6

95 percent confidence interval:

-Inf 98.35577

sample estimates:

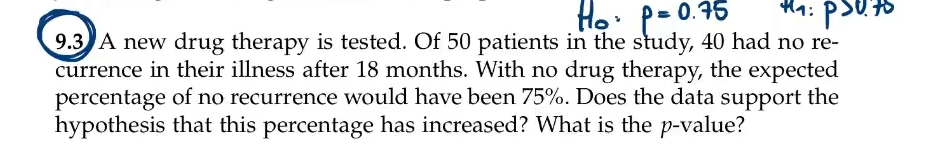
mean of x

98.24923

0.05>1.205e-07? sì! 0.01>? sì! rifiuto ipotesi nulla in favore dell’alternativa

Less-> mi dice che la nulla è 98.6, ma si pensa che possa essere 98.2 (98.2<98.6)->less

-Inf 98.35577 sono tutti valori plausibili, infatti 98.6 non c’è->alternativa



binom.test(c(40,10),alternative = "greater", p = 0.75)

Exact binomial test

data: c(40, 10)

number of successes = 40, number of trials = 50, p-value = 0.2622

alternative hypothesis: true probability of success is greater than 0.75

95 percent confidence interval:

0.6844039 1.0000000

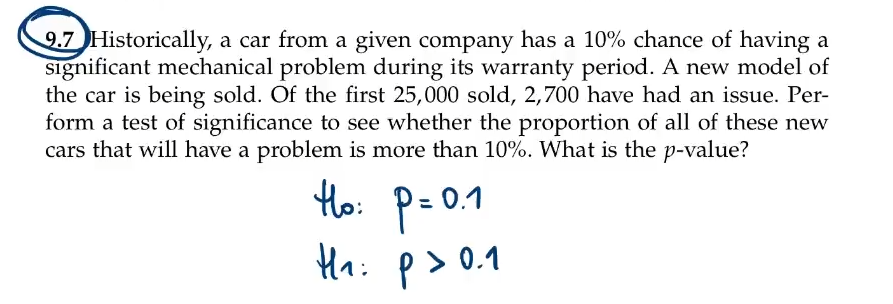
sample estimates:

probability of success

0.8

0.05>0.2622? no! 0.01>? no! Rifiuto l’ipotesi alternativa in favore della nulla

Greater->”percentage has increased?”



binom.test(c(2700,(25000-2700)),alternative = "greater", p = 0.1)

Exact binomial test

data: c(2700, (25000 - 2700))

number of successes = 2700, number of trials = 25000, p-value = 1.588e-05

alternative hypothesis: true probability of success is greater than 0.1

95 percent confidence interval:

0.1047849 1.0000000

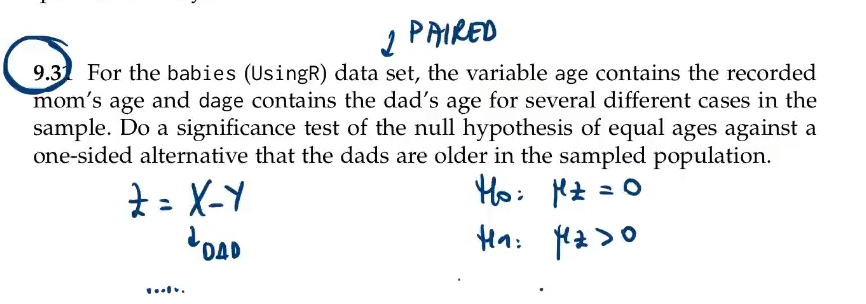
sample estimates:

probability of success

0.108

0.05>1.588e-05? Sì! 0.01>? Sì! Rifiuto l’ipotesi nulla a favore dell’alternativa

Greater->”more than”



t.test(babies$dage, babies$age, alternative = "greater", mu = 0, paired = TRUE)

Paired t-test

data: babies$dage and babies$age

t = 17.392, df = 1235, p-value < 2.2e-16

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

3.047148 Inf

sample estimates:

mean of the differences

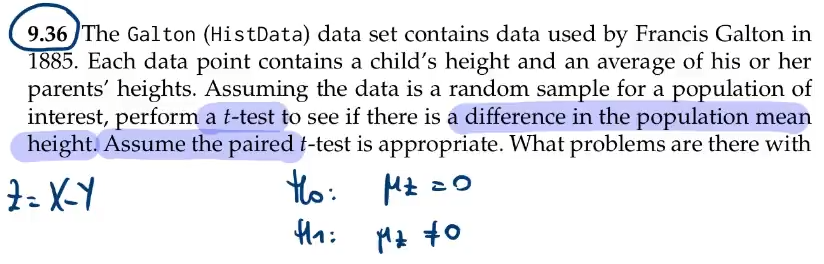
3.365696

Paired=TRUE!->si riferiscono entrambe ad un bambino. Misuro il bimbo e dal bimbo ricavo età mamma e papà

Motivo per cui uso due variabili (babies$dage, babies$age)

Greater->età papà>età mamma

0.05> 2.2e-16? Sì! 0.01> 2.2e-16? Sì->abbandonare ipotesi nulla a favore dell’alternativa



t.test(Galton$child, Galton$parent, alternative = "two.sided", mu = 0, paired = TRUE)

Paired t-test

data: Galton$child and Galton$parent

t = -2.8789, df = 927, p-value = 0.004082

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.36949983 -0.06993982

sample estimates:

mean of the differences

-0.2197198

Two.sided->mi interessa la differenza, non se è > o <

Paired=true ->si riferiscono entrambe ad un unico bambino

0.05>0.004082? sì! 0.01>? sì! Abbandonare ipotesi nulla a favore dell’alternativa